Accelerated and interpretable oblique random survival forests

Byron C. Jaeger[1], Sawyer Welden[1], Kristin Lenoir[1], Jaime L. Speiser[1], Matthew W. Segar[2], Ambarish Pandey[3], and Nicholas M. Pajewski[1]

[1]Wake Forest University School of Medicine, Winston-Salem NC
[2]Texas Heart Institute, Houston TX
[3]University of Texas Southwestern Medical Center, Dallas TX

Introduction

- Risk prediction may reduce disease burden by guiding strategies for prevention and treatment.
- Oblique random survival forests (RSFs) have high prediction accuracy, but also have high computational overhead and few methods for interpretation.[1]
- We developed methods to make oblique RSFs faster and more interpretable.

Background

Axis-based and oblique trees

Decision trees grow by splitting a set of training data into two subsets with maximally different expected outcomes. When the new subsets are formed based on *a single predictor*, the tree is **axis-based** because the splits of the data are perpendicular to the axis of the predictor. When subsets are formed based on **a linear combination of predictors**, the tree is **oblique** because the splits are neither parallel nor perpendicular to the axis.

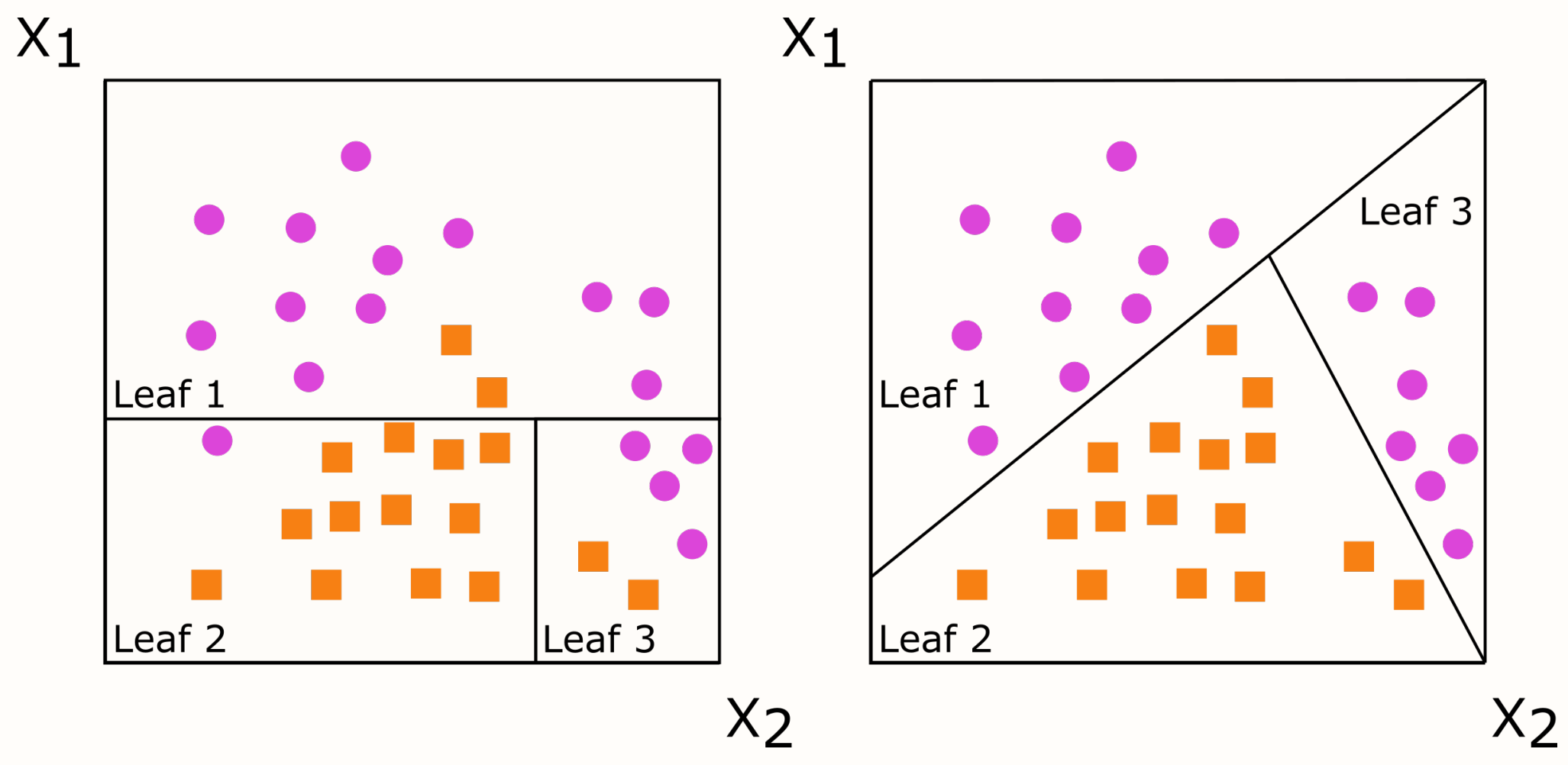


Figure 1: Decision trees with axis-based splitting (left) and oblique splitting (right). The oblique trees do a better job of separating the two classes.

Problems:

- Oblique RSFs have high computational overhead due to the large number of possible oblique splits to search through.
- Oblique RSFs are difficult to interpret as there are few methods to estimate predictor importance using an oblique RSF.

Standard technique for predictor importance: permutation

Definition: For predictor X, the increase in prediction error after X is randomly permuted.
Problem: Permuting a predictor does not account for the size of coefficients attached to it in linear combinations. In oblique random forests, important predictors are likely to have larger coefficients, and ignoring this leads to unreliable importance estimates.

Accelerating the oblique random survival forest

Partial Newton Raphson scoring

We propose to identify linear combinations of predictors in non-leaf nodes by applying Newton Raphson scoring to the partial Cox likelihood.
Hypothesis: Using one Newton Raphson iteration to identify linear combinations of predictors will yield an efficient oblique RSF with no loss of prediction accuracy compared to other learners:

- `aorsf-fast`: One Newton Raphson iteration
- `aorsf-cph`: Newton Raphson until convergence (*i.e.*, fits a Cox proportional hazards model)
- `obliqueRSF`: Penalized regression (*i.e.*, the original method used for oblique RSFs)
- Many other learners (see ArXiv paper for full description - QR code is in the references)

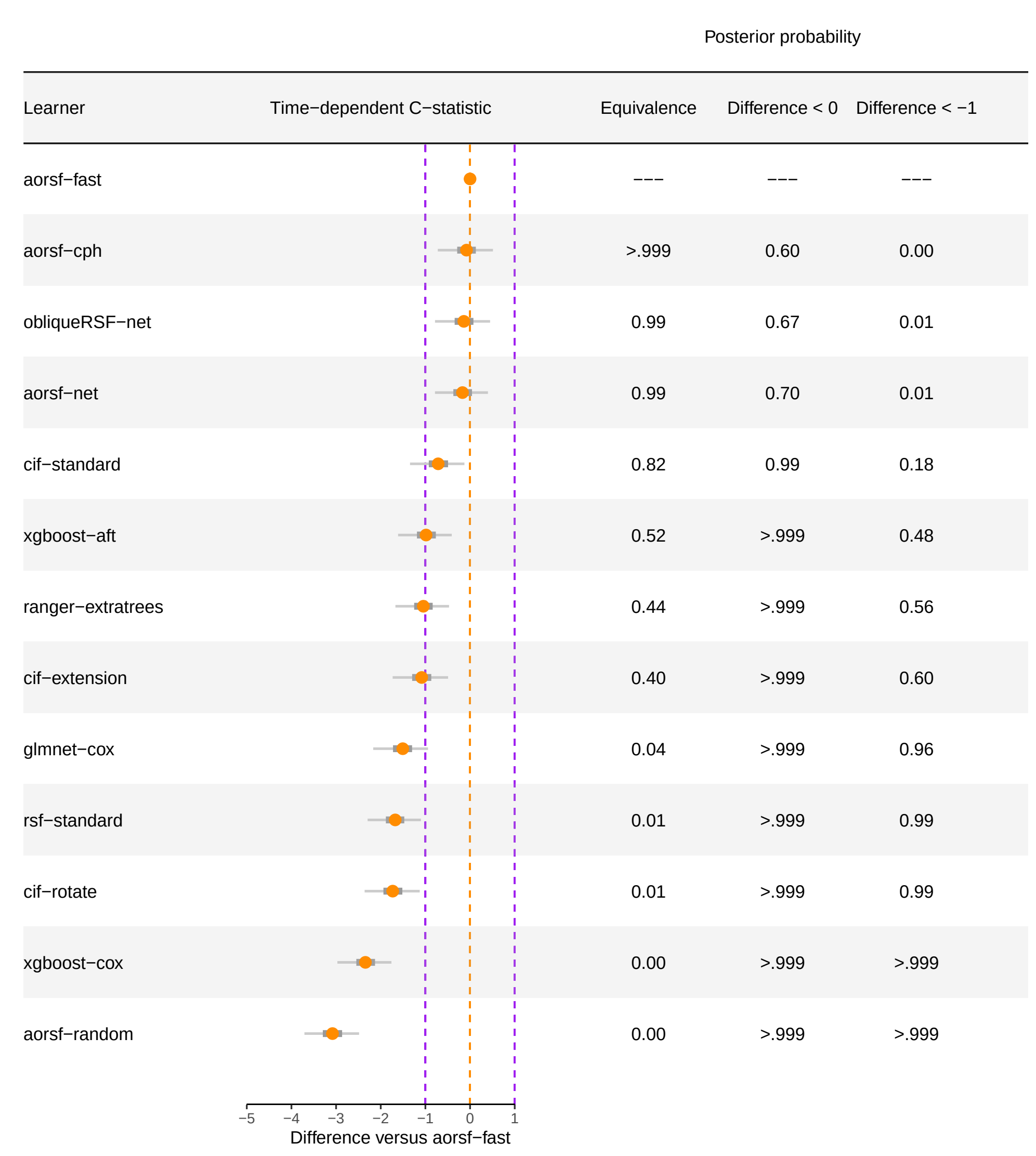
Introducing negation importance

Definition: For predictor X, the increase in an oblique random forest's prediction error after coefficients attached to X are multiplied by -1.
Hypothesis: negation importance will improve detection of signal versus noise predictors.

Benchmarks

Prediction accuracy

We evaluated aorsf-fast's Discrimination (C-statistic) compared to several machine learning algorithms in 35 distinct risk prediction tasks drawn from 21 distinct datasets. Using Bayesian mixed models, inferences on equivalence and inferiority of aorsf-fast versus other learners were made.



		Posterior probability		
Learner	Time−dependent C−statistic	Equivalence	Difference < 0	Difference < −1
aorsf−fast		–––	–––	–––
aorsf−cph		>.999	0.60	0.00
obliqueRSF−net		0.99	0.67	0.01
aorsf−net		0.99	0.70	0.01
cif−standard		0.82	0.99	0.18
xgboost−aft		0.52	>.999	0.48
ranger−extratrees		0.44	>.999	0.56
cif−extension		0.40	>.999	0.60
glmnet−cox		0.04	>.999	0.96
rsf−standard		0.01	>.999	0.99
cif−rotate		0.01	>.999	0.99
xgboost−cox		0.00	>.999	>.999
aorsf−random		0.00	>.999	>.999

Figure 2: Expected differences in C-statistic between aorsf-fast and other learners. A region of practical equivalence is shown by purple lines, and a boundary of non-zero difference is shown by an orange line.

Predictor importance

We evaluated negation importance and several other methods based on how well they discriminated between relevant and irrelevant predictors in a simulation study. In addition to negation, we considered

- ANOVA importance: the proportion of times a given predictor's p-value is <0.001 in linear combinations
- Permutation importance: increase in prediction error after a given predictor is randomly permuted.
- Shapley importance: the expected absolute contribution of a predictor to a model's predictions.

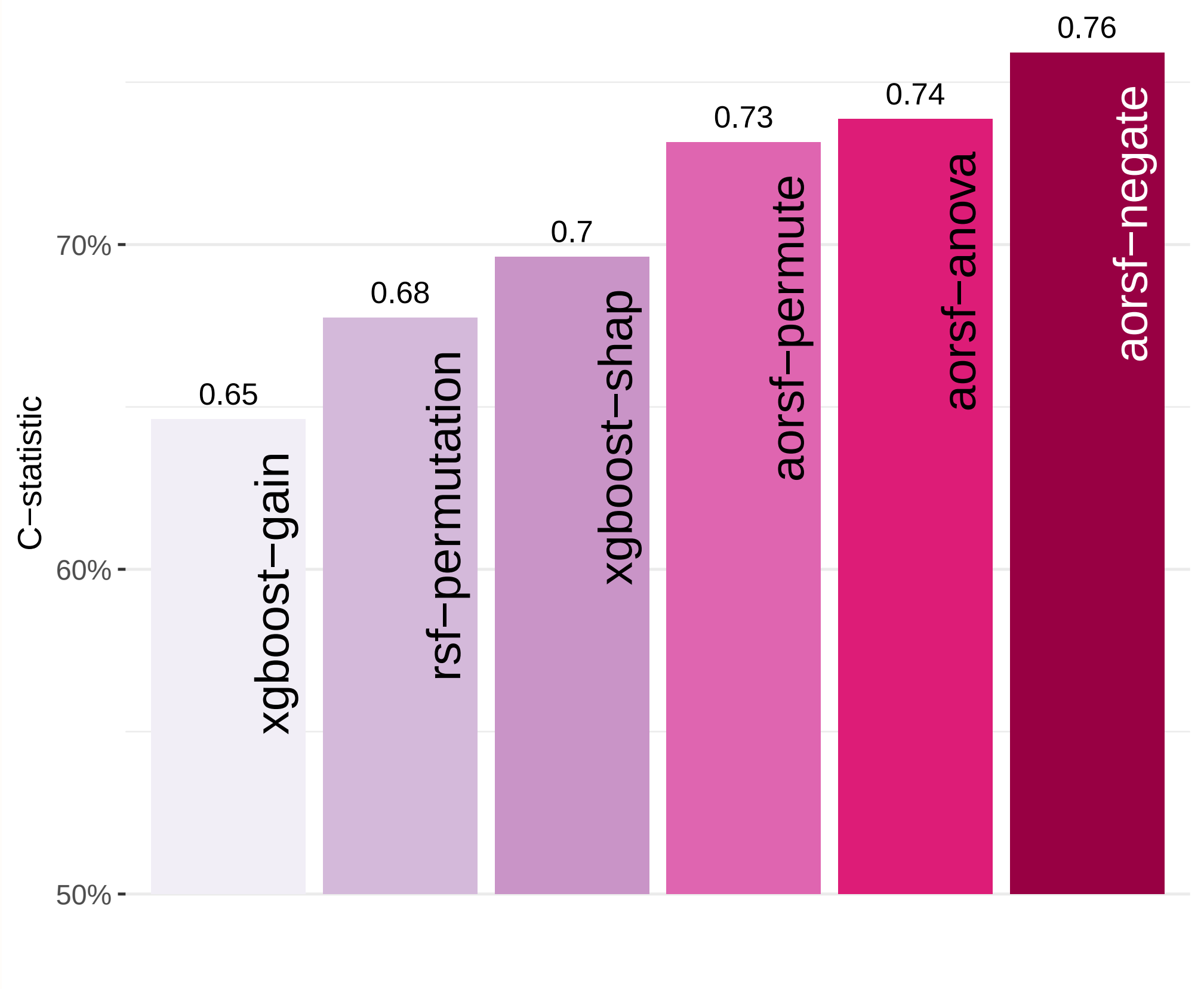


Figure 3: C-statistic for variable selection using different techniques to estimate predictor importance.

Computational efficiency

We evaluated the computational efficiency of the aorsf-fast compared to other machine learning algorithms.

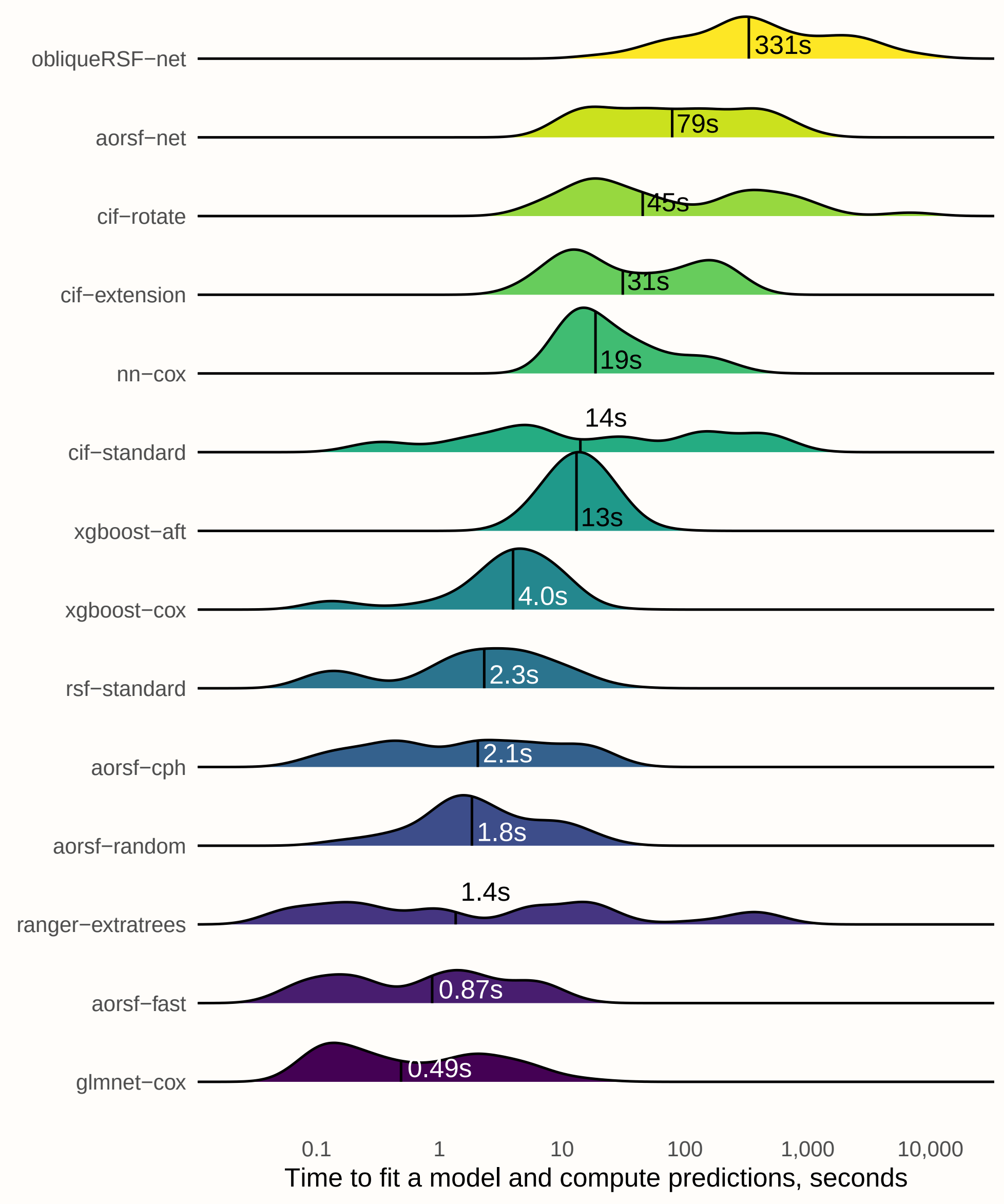


Figure 4: Distribution of time taken to fit a prediction model and compute predicted risk. The median time, in seconds, is printed for each learner

Summary and conclusions

- Oblique RSFs have exceptional prediction accuracy.
- We have developed a method to fit oblique RSFs efficiently, with no loss of prediction accuracy.
- We have also introduced a general method to estimate predictor importance with oblique random forests (not just RSFs), and demonstrated its effectiveness specifically for the oblique RSF.

Acknowledgments

Research reported in this publication was supported by the Center for Biomedical Informatics, Wake Forest University School of Medicine. The project described was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number UL1TR001420. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

aorsf website

aorsf paper

[1] Jaeger, Byron C., et al. "Oblique random survival forests." *The Annals of Applied Statistics* 13.3 (2019): 1847-1883.

Wake Forest Biomedical Informatics Colloquium 2022